

Networking Archives: Quantitative History and the Contingent Archive

Yann Ryan^a, Sebastian Ahnert^b and Ruth Ahnert^a

^aQueen Mary, University of London, Mile End Road, London E1 6NS

^bUniversity of Cambridge, Cambridge CB2 1TN

Abstract

Recent years have seen a growth in the use of network analysis on large datasets of correspondence, but studies of the epistemological basis for findings have not seen a commensurate increase. The latter are important because although large, these datasets can only ever represent a fraction of the total available correspondence, and most historically contingent letter archives have significant amounts of missing or uncertain data or records. This paper outlines three approaches to the study of missing network data: first, we suggest some strategies for dealing with missing data, beginning with understanding in detail the type and extent of missing data, second, we outline a method for understanding the effect that missing data has specifically on historical letter archives, which compares rank correlations of metrics between the full network and progressively smaller random sub-samples. The experiments show that the most basic metric of network structure, degree, is remarkably robust to random letter removal even when large samples of letters have been removed. Last, the paper argues that the combinatorial effect of joined-up letter networks can be used to further the understanding of the structure of seventeenth-century letter networks and intellectual exchange.

Keywords

network analysis, archival history, missing data, state paper office, seventeenth century

1. Introduction

‘Archives are the ... laboratories of the historian’ [19, p. 9].

Alexandra Walsham’s metaphor opens her history of early modern archives, gesturing to the way that archives both supply and shape the narratives that historians have written. Archives have become laboratories in another way too, rendering us with the metadata and digitised collections for computational experiments. However, as Walsham rightly notes, scholars have ‘rarely paused to consider how and why these repositories came into being, despite the fact that these processes have fundamentally shaped and coloured our knowledge of the past’ [19, p. 9]. This is as true of the digital historian as it is of the analogue: digital scholarship is good at grappling with the relationship between the part and the whole when the whole is the digital collection or corpora with which they are working, but there remains relatively little work on how representative those archives and collections are of the larger contexts within which they sit. Andrew Piper argues that ‘where the social sciences often speak in terms of

CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands

✉ y.c.y.ryan@qmul.ac.uk (Y. Ryan); sea31@cam.ac.uk (S. Ahnert); r.r.ahnert@qmul.ac.uk (R. Ahnert)


🌐 <https://www.phy.cam.ac.uk/directory/ahnerts> (S. Ahnert);

<https://www.qmul.ac.uk/sed/staff/ahnertr.html> (R. Ahnert)

📄 0000-0002-0877-7063 (Y. Ryan); 0000-0002-8503-1580 (R. Ahnert)

© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

“samples” and “bias”, the notion of “representativeness” suggests there is not ultimately some stable, knowable whole against which to limit one’s “bias”’ [13, p. 9].

Whilst as a field we have moved past the claims that large data sets give us any kind of comprehensiveness (see [8, pp. 7–8] and rebuttal [2]), there is still little work that shows how the models developed in the field of computational humanities are affected by the historically specific and contingent histories of the archives they leverage. This paper seeks to demonstrate ways that the contours of archives might be quantitatively described in order to understand the impact of those contingent histories on computational analysis of that archive at scale, and applies the methods to one such union of multiple large collections of correspondence. We suggest that these quantitative measures may be useful for others working with similarly large yet idiosyncratic network datasets.

1.1. Networking Archives - a case study

The project from which this work arises, ‘Networking Archives’, seeks to reconstruct epistolary networks by bringing together metadata from multiple early modern archives and catalogues to form a meta-archive of ~450,000 letters. The analysis and techniques described in this paper are a work in progress from a larger project: outputs will include a monograph as well the code used for this analysis and others. The letter, by design, is a technology of dispersal [6, p. 9]. The efforts to re-assemble the oeuvre of specific letter writers or communities is behind the tradition of the ‘collected correspondence’, edited volumes that reunite the incoming and outgoing missives of a given individual or individuals. Online, fully-searchable repositories of correspondence text and metadata, such as *The Electronic Enlightenment* and the *Epsilon project* bring with them the potential for researchers to aggregate correspondence data at scale, in order to begin the task of reconstructing communities of knowledge. The model of aggregation can be extrapolated, to include ever larger and larger bodies of letters. But the same problem remains, that we are dealing with overlapping partial worldviews. The question is how we can model the impact of that partiality on the knowledge we can reconstruct. In this paper, intended as a report on work-in-progress, we share the ways we are dealing with and acknowledging partial and perspectival data.

The networks of correspondence which form the basis of this project come from two digital sources: Early Modern Letters Online (EMLO) and a curated, cleaned dataset derived from State Papers Online (SPO).¹ EMLO is a union of more than a hundred individual catalogues of correspondence, including a large core of metadata from digitised versions of the Bodleian card catalogue (BCC hereafter). This card catalogue was the product of work by two twentieth century employees and one volunteer in the Bodleian and ultimately based on individual and idiosyncratic acquisitions by the Library over time, resulting in an ‘ad hoc’ and ‘iterative’ set of metadata [10]. EMLO has been formed for the purpose of understanding a particular phenomenon—the Republic of Letters—and as such makes no claims to ‘representativeness’ in terms of a more general European network of information. Instead, EMLO is best described as a curated collection based around the correspondence networks of three key groups: a group of early seventeenth-century scholars known as the ‘Dutch Humanists’, a loosely-defined ‘circle’ of correspondents with the mid seventeenth-century polymath Samuel Hartlib at its centre, and the early members of the Royal Society from the 1660s onwards.

¹This data preparation builds on tools developed by the team members for previous projects, reported on in Hyvönen et al. (2019). Analysis was carried out on a version of the data from late 2018.

The correspondence in ‘State Papers Online’ (SPO), which is the digitised papers of the State Paper Office, has cohered as the result of a very different set of processes. The State Paper Office was established in 1610 with the aim of collecting the English state’s private and working manuscripts in a single place, and was to become the principal archive and working library for the parliamentary executive, while essentially being the private papers of the monarch. We might therefore expect this ‘official’ record of the English state to present a more unified or coherent worldview than EMLO. In fact, the State Papers are also full of partial or shifting perspectives: individual secretaries often viewed their official documents as ‘private’ and kept them as their possessions on leaving office. Some, such as the papers of secretary Morrice disappeared almost entirely, while those of Coventry with further private family papers added in were only returned much later.² Document collection in the early years was often chaotic and piecemeal (Marshall, n.d.). Thus, in March 1669 Charles II issued a warrant that all officers of state were to allow Joseph Williamson, the Keeper of State Papers ‘to peruse all such records & memorialls as now are in your custody, & from them to make Transcripts of whatsoever Treatyes, Leagues, Commissions conducing thereto, & publike Grants which he [...] shall deeme fit for Our Service’ [14, p. 188]. Records swelled in consequence thereafter. Furthermore, the metadata we have available to us is derived from nineteenth-century printed calendars, which do not include as such the Northern department of the Office of the secretaries of state as it existed from the Restoration onwards, although confusingly some other papers of the Northern secretaries *are* included.

1.2. Evaluating the Contours of Multiple Archives

As a record of social interactions between discrete entities, these datasets are naturally suited to analysis using tools from network science—a field concerned with understanding the structure and dynamics of complex networks as a series of mathematical phenomena. While EMLO and SPO have very different histories—one being a modern union catalogue and the other the product of a long and complex institutional history—the elements of contingency that have shaped them result in very similar network topologies. Both contain a small number of highly connected nodes (the entities or points to be connected in a network graph), and a large number of poorly connected nodes. Moreover there exists a continuum of connectivity between these two extremes that renders these networks ‘scale-free’, meaning that in every subregion of the network, on almost every scale, we have a few relatively well connected nodes and many relatively poorly connected ones. A scale-free network can be recognised by plotting the distribution of the number of connections for each node - its ‘degree’ - which in the case of scale-free networks follows a straight line on a double logarithmic plot of this kind (figure 1).

So what underlies this broad structural similarity, and how does it play out across the different dimensions of our data? To tease this apart we have analysed the contours of our respective archives along five dimensions: size-distribution, space, time, absence, and interconnectedness. Each of these has helped reveal the way that the respective histories of the archive can impact analyses. We do not have space to go into all five of these here, but aspects of size-distribution, absence and interconnectedness are particularly instructive in explaining the structure observed in figure 1.

The reason for the topology of the networks that we see in figure 1 above can be gleaned in large part by understanding the letter distribution across its constituent parts. In the case of

²For example the Conway Papers were removed from the State Papers office by Secretary Conway while he was still in office, and only returned by their then owner, John Croker, in 1857.

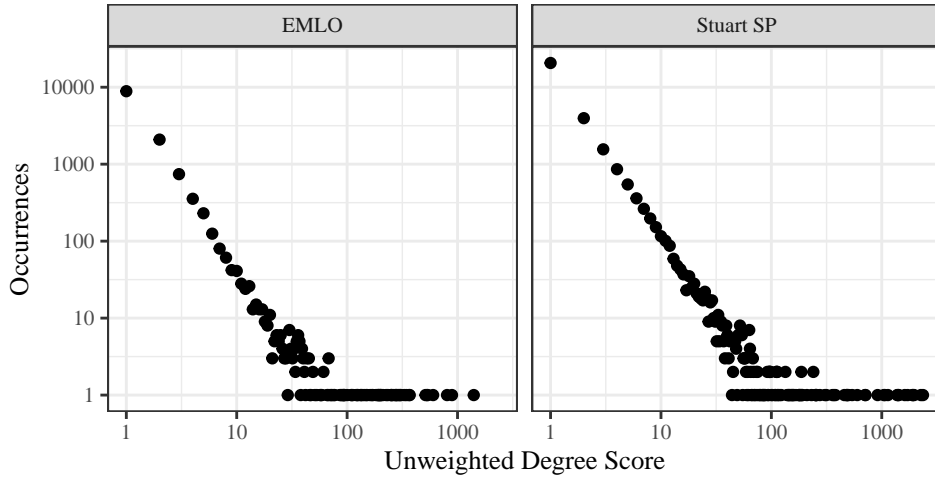


Figure 1: Degree distribution of EMLO and State Papers Networks, plotted on a double logarithmic scale.

EMLO that means understanding the number of letters per catalogue. It is made up of about 110 constituent catalogues, and the distribution of the number of letters across these catalogues is heavily skewed with 80% of the letters coming from the top twenty-three catalogues (21%). Moreover the nature of these respective catalogues has a large impact on the contours of those constitutive parts: a correspondence collection usually, though not always, prioritises an individual at its centre. While most real-world networks tend to cohere around a series of key hubs, this phenomenon is exaggerated in the EMLO data because it is in effect sampled from the perspective of the hubs, which is not dissimilar to the way that many traditional datasets used for Social Network Analysis (SNA) are based on surveys or responses to a questionnaire which relies on participants listing their connections [see 20].

Figure 1 shows that the degree distribution of the nodes in the State Papers Stuart network is strikingly similar to that derived from EMLO’s data. We could also describe this network as in effect a collection of ego networks (the name for a network derived from the perspective of a single node), as it is in essence the worldview of the principal secretaries of state plus a limited set of other nodes. Other analysis has shown that networks based on earlier Tudor correspondence had similar degree distributions, as most of the connections are assigned to the small number of individuals who formed the ‘gravitational centre’ of the English state [1, p. 32].

The topology however is also a product of missing data. We have found that the patterns of this missing data vary not only between archives, but also between catalogues within EMLO. Contributors to EMLO have collated correspondence inventories based on contemporary or modern print editions; born-digital editions or listings; library catalogues; and/or individual or project archival research, so therefore the metadata available to us varies in type and quality, despite extensive standardisation and reconciliation. The missing information is systematic rather than random, and there is a strong correlation between missing data in some pairs of fields: records missing recipients are more likely to be missing authors, for example.

The Stuart SPO also has missing data but it is not always directly comparable: no data on letter destination has been recorded in SPO, and there is very little truly ‘missing’ author or recipient data in SPO, though there are numerous inferred, ambiguous or unidentified person

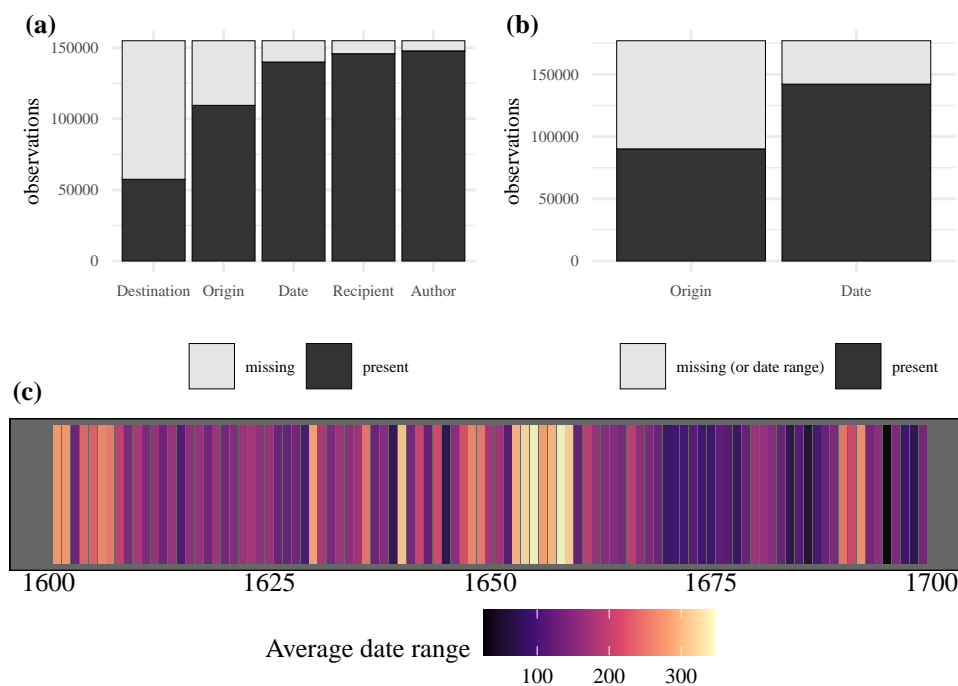


Figure 2: Missing data in EMLO and SPO (Stuart). (a) shows the missing portions of key fields in EMLO. (b) shows missing sections of the origin field, and the section of SPO where the date is given as a range rather than a specific date. Figure (c) shows this date range data further broken down: ‘warmer’ colours indicate a larger mean range for that year.

text strings. Most missing or unknown dates have been given a date range, sometimes spanning the lifetime of the correspondents in the letter, or a period of rule. As an example of the granularity with which we are approaching missing or uncertain data: of the approx. 177,000 records in the Stuart SPO data, 34,829 have an estimated date range rather than a single date (figure 2 (b)). 33,386 of these have a date range of one year or less. The majority of these (21,311) have a date range of exactly 10 days (11 days after 1700) because of uncertainty between Julian and Gregorian calendars. This can be visualised with years represented by vertical bars, and the mean date range by colour, with ‘hotter’ colours indicating less accurate dates (figure 2 (c)). This plot shows a progression from less to more accurate data over the course of the seventeenth century.

This fine-grained data on the ‘known’ partiality of our networks has helped us add necessary caveats or controls to our results: for example, knowing that the SPO and BCC data usually does not contain information on place of destination means that our analysis of the geographic origins and transmission of intelligence or information takes that into account. Similarly, temporal analysis of the formation of these networks must allow for the fact that early records are more likely to be missing either a sender or recipient.

1.3. Missing Data and Network Analysis

Recent years have seen a number of projects in the computational humanities using network theory as the basis for historical analysis [1, 4, 18]. Many of these are based on incomplete data. Because historians are used to working with partial archives, there is some well-founded scepticism about the impact of such absence on large-scale computational analysis. However,

in network science and cognate fields, there is a more established discourse and methodology around the nature and likely impact of incomplete data. For example, missing social network analysis data can be the result of boundary specification problems (difficulties with establishing the parameters of the network) or data collection inaccuracy (incomplete or biased surveys, for example). Studies have found that these types of missing data are likely to result in individual missing nodes or edges [9, p. 248]. Our data, by comparison, is more likely to have missing hub nodes along with most of their connections, because entire personal archives have been lost, or are currently unavailable. As the figures above have shown, data derived from archives like ours are also likely to be missing or have inaccurate years or year ranges because longitudinal historical data can be interrupted by war, change in executive, or bureaucratic practices. However, there is little understanding of the impact of this missing data on network results. Intuitively, one might assume that network metrics based on partial archives are therefore unreliable. It is possible, however, to quantitatively test this assumption.

Researchers have previously used quantitative methods to estimate the effect of missing nodes and edges on network measures [16, 17]. Costenbader and Valente, for example, analysed the sensitivity of eleven centrality measures using social networks derived from questionnaire results and found that eigenvector centrality (a recursive measure of a node’s importance, based on the importance of its connections) was particularly robust as a measure of centrality using sampled data [3, p. 299]. To understand the effect that missing network information might have on our network, we adapted a technique developed by Matthew Peeples for measuring the sensitivity of network measures to random node removal from archaeological networks [12]. This technique involves randomly removing progressively larger random samples from the network, and comparing the results to those found in the full network. Using this, we can infer how various types of gaps in the data might affect the rankings of individual nodes (figure 3). The code for this, as well as a version with a user-friendly interface, will be made available on the project’s Github repository to coincide with a publication publishing full results.

What we have found is that even with very large amounts of letters removed, for those remaining nodes, there remains a strong correlation between the original ranked score of a node and its rank in the random subsample of the network, and very little variation between random samples. The model shows that individuals who are not removed are very likely to remain in similar ranked positions for degree scores despite significant missing data. Degree is a key measurement of a node’s centrality to a network and we surmise that network analysis of correspondence in large historical collections yields results that are remarkably robust to missing data, likely because we have multiple letters—evidence of links—for many of the node relationships. While caution must still be exercised, because in some cases *none* of the possible links between two individuals will have survived, but it does give us confidence about the use of this most basic measure of network centrality in historical findings.

This robustness is partially because of another phenomenon that we have found to be the result of working with many combined catalogues: the emergence of secondary ‘informal’ catalogues, found at the intersections between the primary ego networks. Like the measure of betweenness centrality— a metric which gives a score to each node based on the number of times it is traversed on the shortest paths between every other pair of nodes and which is often used to find individuals with structural importance but not necessarily the highest total connections—looking at overlapping or intersecting neighbourhoods helps us to find individuals who bridge multiple overlapping ego networks, whose significance was harder to derive before the creation of union catalogues and meta-archives such as EMLO.

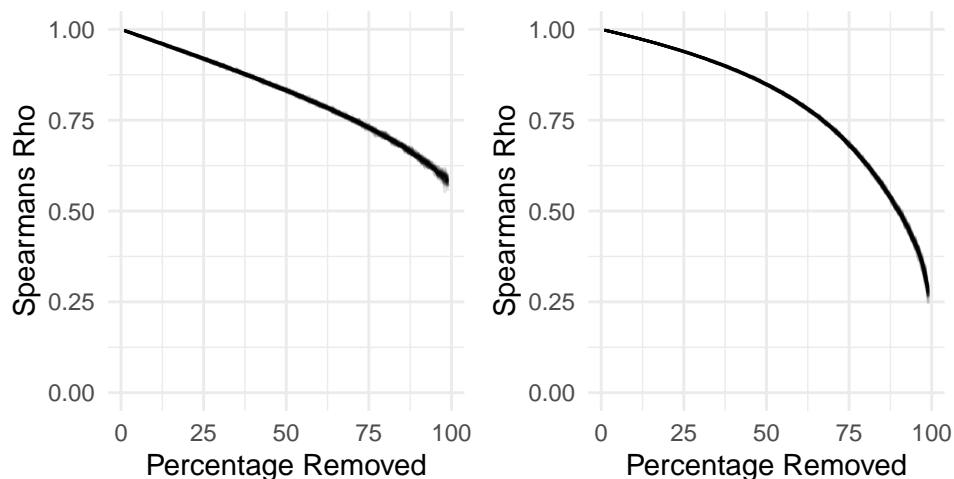


Figure 3: Effect of random letter removal on degree ranks. In the left-hand plot, we removed progressively larger random samples of letters from the full network, recalculated degree ranks and then plotted the resulting Spearman’s rank correlation coefficient between the original and sampled network (this process was repeated 50 times). The right-hand plot is the result when removing edges from a random Barabasi-Albert network with the same number of nodes and edges, which follows a more severe pattern of decay in correlations as larger numbers of edges are removed. What this tells us is that degree rank correlations are remarkably robust to random letter removals from the network.

1.4. Intersections

One of the key aims of the Networking Archives project is to examine the extent to which the multiple overlapping ego networks of the EMLO archive are intermeshed internally, and to what extent they in turn overlap with the political archive of SPO. The reconciliation of EMLO and SPO is still being completed so the following results are based on EMLO only.

The purpose for which EMLO was originally created means that the catalogues of correspondence are deeply enmeshed and overlapping. Even when the individuals at the centre of catalogues are not directly connected by an edge, it is likely they will have correspondents in common. We exploit this fact to uncover or highlight ‘emergent’ ego networks of individuals whose correspondence has not been consolidated into a single collection. To do this, we developed a tool that finds the intersecting set of correspondents for each pair of individuals in the network—whether they corresponded with each other or not (figure 4). We have found that in partial networks, individuals with a high number of these ‘shared correspondents’ are often important figures whose contributions to their particular networks have been overlooked. To find these emergent informal catalogues, we looked for individuals who a) had many significant intersecting correspondents, b) did not have their own catalogue or edited collection and c) could be found across numerous catalogues (table 1)

From these criteria, the seventeenth-century Scottish minister John Dury emerged as an interesting case study. Despite his prominent work as a pamphlet writer, minister, diplomat, tutor, and theologian, there is no single ‘Dury Archive’: John Dury left many of his papers with Samuel Hartlib when he travelled on the continent, and therefore the largest single collection of Dury material can be found within the Hartlib papers, held by Sheffield University. Because of this, he does not have a centralised collection of correspondence in EMLO, although parts of it have been collected in printed editions elsewhere [5, p. 39]. Only a small number (1,372) of

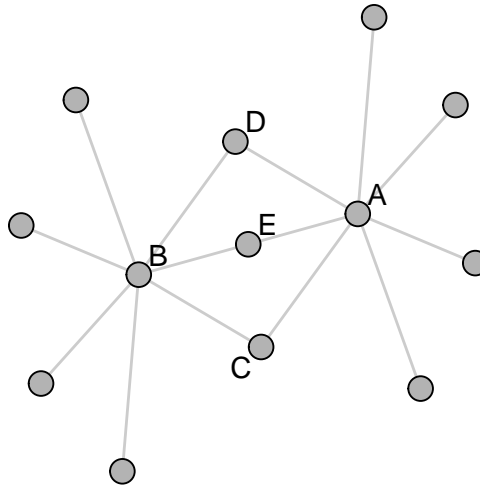


Figure 4: The Overlaps Tool lists the overlapping set (C, D, E) of neighbours of A and B for each pair of nodes in the network—whether or not A and B are also direct neighbours.

Dury’s total surviving letters are listed in EMLO, spread across a number of catalogues, and therefore a ‘standard’ quantitative analysis of Dury’s correspondence might underestimate his importance to the network of which he forms a part.

With a union catalogue like EMLO we can infer the centrality of individuals even without a centralised catalogue. As well as appearing as an author and recipient in Hartlib’s catalogue, Dury can be found in eight others on EMLO, albeit in much smaller numbers. In this way Dury’s centrality is no longer defined by his appearance in a single catalogue, but by his own centrality, inferred from his appearance across other collections, for example the three letters found in a printed edition of the letters of Joseph Mede, or the two ‘lost letters’ of his to Robert Boyle found in a list compiled by William Wotton shortly after Dury’s death [11, pp. 804, 865–65, 866–67], [7, p. 246]. Like scientists inferring the presence of black holes by their absence and the behaviour of objects around them, looking at individuals and their overlaps allows us to establish a sense of their contribution to the network, even when their correspondence has not been systematically collected in one place.

For Dury, this matches up with what we know about him as a figure within a European intellectual network. A prominent Irenicist (working towards a peaceful Protestant union of churches), Dury was in regular contact with theologians, scholars and diplomats throughout Europe. Dury spent much of his life on the move: living or travelling through the Netherlands, Sweden, Poland, Germany, and Switzerland, communicating state affairs to diplomats in England and moving on when he had established contacts, aiming to continue the correspondence by letter after he had left. His mobility and continental connections meant that he acted as a ‘bridge’ in a European network of communication centred around Samuel Hartlib, as well sending intelligence to Cromwell’s secretary of state, John Thurloe, in the mid-1650s. Finding Dury amongst established archives, and specifically linking evidence of his correspondence across multiple catalogues, we can more fully exploit the surviving records of his contributions, even though most of his correspondence is either lost or not available as structured metadata.

Table 1

Individuals in EMLO, arranged by number of times they share at least two correspondents with another individual. Individuals without their own catalogue are highlighted in gray

Name	Overlaps	Degree	Catalogues
Oldenburg, Henry	734	526	17
Wallis, John (Dr)	678	389	10
Huygens, Constantijn	618	1415	17
Vossius, Gerardus Joannes	588	925	12
Boyle, Robert	561	511	13
Sancroft, William	527	927	2
Dury, John	515	283	9
Smith, Thomas	501	369	3
Mersenne, Marin	489	254	15
Komenský, Jan Amos	474	234	9
Huygens, Christiaan	472	372	15
Hartlib, Samuel	469	364	10
Vossius, Isaac (Dr)	456	313	8
Charles II, King of England, Scotland, and Ireland	447	37	4
Groot, Hugo de	441	540	13
Charlett, Arthur (Reverend)	436	441	4
Saumaise, Claude de	415	34	11
Polyander van den Kerckhoven, Johannes	407	177	5
Lister, Martin	402	245	7
Hevelius, Johannes	399	36	9

1.5. Conclusion

The growth of the use of network theory in the humanities over the past decade has been accompanied by self-reflection on its use: at one extreme, it has been used to make *longue duree* claims about hugely complex historical forces [15], while at the same time other researchers have, quite rightly, pointed out that networks based on assembled ego networks may tell us more about the collection practices than about a historical phenomenon more generally (Weingart 2011). This paper has put forward some techniques for understanding and working with partial network data as found in historically contingent archives, in a way which we believe strengthens the epistemological underpinnings of historical network analysis.

Some of the unknown data is ‘known’: partially missing people, dates or place names which we can measure, understand and visualise using the methods as described above. The much larger part of the missing data is truly unknown: destroyed, intercepted or unavailable archives, not to mention the part of the network reliant on oral exchange. Despite this, our findings suggest that network rankings stay remarkably stable when large parts of the network are removed, either at random or by removing entire catalogues—we can also infer that this would be the case were additional catalogues *added* to a meta-collection.

Crucially, the method allows us to reconstruct likely connections in incomplete, historically contingent collections. Quantitative studies of the Republic of Letters have so far focused on individuals whose collections have survived in a single archive or have been assembled

afterwards in a single repository.³ This may have the effect of biasing results towards letters received by geographically stable nodes. John Dury’s centrality to the network, on the other hand, was one based on breadth rather than overall strength, and is harder to measure because of the geographic dispersal of his letters: establishing network strengths across catalogues helps to find these intersecting nodes, and counteract the natural bias towards histories based on individuals as found in formal collections or nineteenth-century printed editions.

The adoption of linked open data means that access to linked discrete archives is set to increase, and existing datasets will continue to expand. The ‘Networking Archives’ project, for example, will link together three large datasets, and one of these, EMLO, is continually expanding its metadata with new catalogues of correspondence. Our understanding of John Dury is a direct benefit of this: almost seventy additional letters by John Dury found in the *Stuart State Papers Online* will, when linked to the existing metadata, help the project to understand more fully his role as a provider of intelligence to the English state. This presents an enormous opportunity: each new addition leads to further intersections and helps us to understand both the added network and the entire network as a whole. This work is part of a larger project, from which will come a project volume, code and analysis. Through this we hope to be transparent about the ways in which our data is biased or partial, but also reassure readers and other historical network analysis practitioners that results do have validity, despite missing data.

Author contributions

YR conceived original idea; adapted code; wrote and revised manuscript; SA wrote and revised manuscript; developed idea; checked code and statistics; RA supervised PI; developed idea; wrote and revised manuscript;

Acknowledgments

This work is funded by the AHRC as part of the project Networking Archives (AH/R014817/1). The work we report on here builds on the work of the project team past and present, including Philip Beeley, Arno Bosse, Howard Hotson (PI), Miranda Lewis, Esther van Raamsdonk, and Matthew Wilcoxson. We would especially like to thank Philip Beeley, Esther van Raamsdonk, Miranda Lewis and Howard Hotson for their comments on this paper, and Matthew Peeples for developing the original robustness code and his permission to adapt it. For a full list of data contributors to Early Modern Letters Online, see http://emlo-portal.bodleian.ox.ac.uk/collections/?page_id=2259.

References

- [1] R. Ahnert and S. E. Ahnert. “Metadata, Surveillance and the Tudor State”. In: *History Workshop Journal* 87 (Jan. 2019), pp. 27–51. ISSN: 1363-3554.

³The *Mapping the Republic of Letters* project, for example, has taken as case studies Athanasius Kircher and John Locke. The bulk of the former’s correspondence is preserved in the Archives of the Pontifical Gregorian University in Rome, and the latter’s was assembled as a multi-volume scholarly edition by Oxford University Press, and made available digitally on the Electronic Enlightenment website.

- [2] K. Bode. “The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”. In: *Modern Language Quarterly* 78.1 (Mar. 2017), pp. 77–106. ISSN: 0026-7929.
- [3] E. Costenbader and T. W. Valente. “The stability of centrality measures when networks are sampled”. en. In: *Social Networks* 25.4 (Oct. 2003), pp. 283–307. ISSN: 03788733. (Visited on 07/17/2020).
- [4] C. Edmondson and D. Edelstein, eds. *Networks of enlightenment: digital approaches to the republic of letters*. Oxford University studies in the Enlightenment 2019:06. OCLC: on1057786719. Liverpool: Liverpool University Press on behalf of Voltaire Foundation, 2019. ISBN: 9781786941961.
- [5] M. Greengrass. “Archive Refractions: Hartlib’s Papers and the Workings of an Intelligencer”. In: *Archives of the Scientific Revolution: The Formation and Exchange of Ideas in Seventeenth-Century Europe*. Ed. by M. Hunter. Woodbridge: Boydell Press, 1998, pp. 35–48.
- [6] H. Hotson and T. Wallnig. “Introduction”. In: *Reassembling the Republic of Letters in the digital age: standards, systems, scholarship*. Göttingen: Göttingen University Press, 2019, pp. 7–23. ISBN: 9783863954031.
- [7] M. Hunter, A. Clericuzio, and L. M. Principe, eds. *The Correspondence of Robert Boyle*. London: Pickering and Chatto, 2001.
- [8] M. L. Jockers. *Macroanalysis: digital methods and literary history*. Topics in the digital humanities. Urbana: University of Illinois Press, 2013. ISBN: 9780252037528 9780252079078 9780252094767.
- [9] G. Kossinets. “Effects of missing data in social networks”. en. In: *Social Networks* 28.3 (July 2006), pp. 247–268. ISSN: 03788733. DOI: 10.1016/j.socnet.2005.07.002. (Visited on 06/15/2020).
- [10] M. Lewis. *Ghosts in the Machine: (Re)Constructing the Bodleian’s Index of Literary Correspondence, 1927-1963*. 2013. URL: <http://www.culturesofknowledge.org/?p=295>.
- [11] J. Mede. *The works of the pious and profoundly-learned Joseph Mede, B.D. sometime fellow of Christ’s College in Cambridge*. London: Printed by Roger Norton for Richard Royston, 1677.
- [12] M. Peeples. *Network Science and Statistical Techniques for Dealing with Uncertainties in Archaeological Datasets*. 2017.
- [13] A. Piper. *Enumerations: data and literary study*. London: The University of Chicago Press, 2018. ISBN: 9780226568614 9780226568751.
- [14] M. Riordan. “”The King’s Library of Manuscripts”: The State Paper Office as Archive and Library”. In: *Information & Culture* 48.2 (2013), pp. 181–193. ISSN: 21648034, 21663033.
- [15] M. Schich et al. “A network framework of cultural history”. en. In: *Science* 345.6196 (Aug. 2014), pp. 558–562. ISSN: 0036-8075, 1095-9203. (Visited on 07/17/2020).
- [16] J. A. Smith and J. Moody. “Structural effects of network sampling coverage I: Nodes missing at random”. en. In: *Social Networks* 35.4 (Oct. 2013), pp. 652–668. ISSN: 03788733. (Visited on 06/15/2020).

- [17] J. A. Smith, J. Moody, and J. H. Morgan. “Network sampling coverage II: The effect of non-random missing data on network measurement”. en. In: *Social Networks* 48 (Jan. 2017), pp. 78–99. ISSN: 03788733. (Visited on 06/15/2020).
- [18] I. Van Vugt. “Using Multi-Layered Networks to Disclose Books in the Republic of Letters”. en. In: *Journal of Historical Network Research* 1 (Oct. 2017), pp. 25–51. ISSN: 2535-8863. (Visited on 06/15/2020).
- [19] A. Walsham. “The Social History of the Archive: Record-Keeping in Early Modern Europe”. en. In: *Past & Present* 230.suppl 11 (2016), pp. 9–48. ISSN: 0031-2746, 1477-464X. (Visited on 06/15/2020).
- [20] S. Wasserman and K. Faust. *Social Network Analysis Methods and Applications*. Cambridge: Cambridge University Press, 1994.